American Bar Association
www.supremecourtpreview.org

Nos. 07-1428 & 08-328

IN THE
Supreme Court of the United States

FRANK RICCI, ET AL.,
Petitioners,
v.

JOHN DESTEFANO, ET AL.,
Respondents.

On Writs of Certiorari to the United States Court of Appeals for the Second Circuit

BRIEF OF INDUSTRIAL-ORGANIZATIONAL PSYCHOLOGISTS AS *AMICI CURIAE* IN SUPPORT OF RESPONDENTS

DAVID C. FREDERICK
Counsel of Record
DEREK T. HO
BARRETT C. HESTER
JENNIFER L. PERESIE
KELLOGG, HUBER, HANSEN,
TODD, EVANS & FIGEL, P.L.L.C.
1615 M Street, N.W., Suite 400
Washington, D.C. 20036
(202) 326-7900

Counsel for Industrial-Organizational Psychologists

March 25, 2009

TABLE OF CONTENTS

Page

TABLE OF AUTHORITIES iii

INTEREST OF *AMICI CURIAE* 1

SUMMARY OF ARGUMENT 3

ARGUMENT 5

I. THE 2003 EXAMINATIONS CONTAINED FATAL FLAWS THAT UNDERMINED THEIR ABILITY TO SELECT THE MOST QUALIFIED CANDIDATES 5

A. Proper Validation Of Employment Tests According To Established Standards Is Essential To Ensuring Fair, Merit-Based Selection 6

B. Contrary To Petitioners' Premise, The 2003 NHFD Examinations Could Not Have Been Validated Under Established Principles 8

1. The Test Designer Conceded That the Exams Did Not Attempt To Measure Command Presence, a Critical Job Attribute 10

2. The Weighting of the Multiple-Choice and Oral Interview Portions of the Exams Was Arbitrary and Could Not Be Validated 12

3. Flaws in the Exam-Development Process Contributed to the Lack of Validity Evidence of the NHFD Tests 19

4. The NHFD Tests Could Not Have Been Validated for Strict Rank-Ordering of Candidates 21

II. THE FLAWS IN THE NHFD PROMOTIONAL EXAMS EXACERBATED THEIR ADVERSE IMPACT ON MINORITY CANDIDATES 23

A. Overweighting Of The Written, Multiple-Choice Portion Of The Exams Increased The Adverse Impact On Minority Candidates 24

B. Selection Of Candidates In Strict Rank Order Also Contributed To Adverse Impact 26

III. CURRENT I/O PSYCHOLOGY RESEARCH SUPPORTS THE USE OF PROMOTIONAL ASSESSMENT CENTERS AS A VALID AND LESS DISCRIMINATORY ALTERNATIVE TO TRADITIONAL TESTING METHODS 28

A. History Of The Assessment Center Model 29

B. Assessment Centers Have Demonstrated Validity In The Context Of Firefighter Promotion 30

C. Assessment Centers Have Been Proven To Reduce Adverse Impact On Minorities 32

CONCLUSION 34

TABLE OF AUTHORITIES

Page

CASES

Biondo v. City of Chicago, 382 F.3d 680 (7th Cir. 2004)..... 28

Bridgeport Guardians, Inc. v. City of Bridgeport, 933 F.2d 1140 (2d Cir. 1991) 28

Chicago Firefighters Local 2 v. City of Chicago, 249 F.3d 649 (7th Cir. 2001) 27-28

Chisholm v. United States Postal Serv., 516 F. Supp. 810 (W.D.N.C. 1980), *aff'd in part*, 665 F.2d 482 (4th Cir. 1981) 18

Ensley Branch, NAACP v. Seibels, 31 F.3d 1548 (11th Cir. 1994) 23

Firefighters Inst. for Racial Equality v. City of St. Louis:

549 F.2d 506 (8th Cir. 1977) 10, 12

616 F.2d 350 (8th Cir. 1980) 16, 17, 22, 32

Griggs v. Duke Power Co., 401 U.S. 424 (1971) 6, 29

Guardians Ass'n of New York City Police Dep't v. Civil Serv. Comm'n, 630 F.2d 79 (2d Cir. 1980) 21, 22

Isabel v. City of Memphis:

No. 01-2533 ML/BRE, 2003 WL 23849732 (W.D. Tenn. Feb. 21, 2003), *aff'd*, 404 F.3d 404 (6th Cir. 2005) 13

404 F.3d 404 (6th Cir. 2005) 18

Jones v. New York City Human Res. Admin., 391 F. Supp. 1064 (S.D.N.Y. 1975), *aff'd*, 528 F.2d 696 (2d Cir. 1976) 14

Kelly v. City of New Haven, 881 A.2d 978 (Conn. 2005) 22

Nash v. Consolidated City of Jacksonville, 837 F.2d 1534 (11th Cir. 1988), *vacated and remanded*, 490 U.S. 1103 (1989), *opinion reinstated on remand*, 905 F.2d 355 (11th Cir. 1990) 17

Pina v. City of East Providence, 492 F. Supp. 1240 (D.R.I. 1980) 13, 22

STATUTES, REGULATIONS, AND RULES

Civil Rights Act of 1964, Tit. VII, 42 U.S.C. § 2000e *et seq.* 5, 26, 29, 33

29 C.F.R. pt. 1607 (*Uniform Guidelines on Employee Selection Procedures*) 5, 6, 7, 9, 12, 15, 16, 18, 21, 29

§ 1607.1(A) 6

§ 1607.1(B) 7

§ 1607.3(A) 7

§ 1607.5(B) 7

§ 1607.5(C) 6

§ 1607.5(G) 21

§ 1607.5(H) 18

§ 1607.9 9

§ 1607.14(B) 9

§ 1607.14(B)-(D) 7

§ 1607.14(B)(3) 10

§ 1607.14(C)(4) 10

§ 1607.14(C)(9) 21

Sup. Ct. R. 37.6 1

ADMINISTRATIVE MATERIALS

Adoption of Questions and Answers To Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures, 44 Fed. Reg. 11,996 (1979) 16, 17

OTHER MATERIALS

Herman Aguinis & Erika Harden, *Will Banding Benefit My Organization? An Application of Multi-Attribute Utility Analysis*, *in Test-Score Banding in Human Resource Selection* 193 (Herman Aguinis ed., 2004) 8, 22

American Psychological Association, *Standards for Educational and Psychological Tests* (1999) 6

Winfred Arthur Jr. et al., *A Meta-Analysis of the Criterion-Related Validity of Assessment Center Dimensions*, 56 Personnel Psychol. 125 (2003) 31

Winfred Arthur Jr. et al., *Multiple-Choice and Constructed Response Tests of Ability*, 55 Personnel Psychol. 985 (2002) 24, 25

Walter C. Borman et al., *Personnel Selection*, 48 Ann. Rev. Psychol. 299 (1997)..... 31

David L. Bullins, *Leading in the Gray Area*, Fire Chief (Aug. 10, 2006), *at* http://firechief.com/management/bullins_gray08102006/index.html 15

Wayne F. Cascio et al., *Social and Technical Issues in Staffing Decisions, in Test Score Banding*, *in Human Resource Selection* 7 (Herman Aguinis ed., 2004) 23, 27

Wayne F. Cascio & Herman Aguinis:

Applied Psychology in Human Resource Management (6th ed. 2005).....20, 32

Test Development and Use: New Twists on Old Questions, 44 Human Res. Mgmt. 219 (2005) 18

John F. Coleman, *Incident Management for the Street-Smart Fire Officer* (2d ed. 2008) 11

Vincent Dunn, *Command and Control of Fires and Emergencies* (1999) 11

Robert D. Gatewood & Hubert S. Feild, *Human Resource Selection* (5th ed. 2001) 13, 14

Barbara B. Gaugler et al., *Meta-Analysis of Assessment Center Validity*, 72 J. Applied Psychol. 493 (1987).....31, 32

Gary M. Gebhart et al., *Fire Service Testing in a Litigious Environment: A Case History*, 27 Pub. Personnel Mgmt. 447 (1998) 32-33

Irwin L. Goldstein et al., *An Exploration of the Job Analysis–Content Validity Process*, *in Personnel Selection in Organizations* 3 (Neil Schmitt & Walter C. Borman eds., 1993) 20

Charles D. Hale, *The Assessment Center Handbook for Police and Fire Personnel* (2d ed. 2004).....30, 33

Chaitra M. Hardison & Paul R. Sackett, *Assessment Center Criterion Related Validity: A Meta-Analytic Update* (2004) 31

James R. Huck & Douglas W. Bray, *Management Assessment Center Evaluations and Subsequent Job Performance of White and Black Females*, 29 Personnel Psychol. 13 (1976) 29

Int'l Ass'n of Fire Chiefs et al.:

Fire Officer: Principles and Practice (2006) 15, 29

Fundamentals of Fire Fighter Skills (2004) 20

Anthony Kastros, *Mastering the Fire Service Assessment Center* (2006) 11

Richard Kolomay & Robert Hoff, *Firefighter Rescue & Survival* (2003) 11

Diana E. Krause et al., *Incremental Validity of Assessment Center Ratings Over Cognitive Ability Tests: A Study at the Executive Management Level*, 14 Int'l J. Selection & Assessment 360 (2006)..... 31

Phillip E. Lowry, *A Survey of the Assessment Center Process in the Public Sector*, 25 Pub. Personnel Mgmt. 307 (1996).....15, 30

Lynn Lyons Morris et al., *How to Measure Performance and Use Tests* (1987)..... 13

Matthew Murtagh, *Fire Department Promotional Tests* (1993) 10, 11, 14

James L. Outtz & Daniel A. Newman, *A Theory of Adverse Impact* (manuscript on file with author) (forthcoming in *Adverse Impact: Implications for Organizational Staffing and High Stakes Selection*, 2009) 25

Miguel Roig & Maryellen Reardon, *A Performance Standard for Promotions*, 141 Fire Engineering 49 (1988) 30

Philip L. Roth et al., *Ethnic Group Differences in Cognitive Ability in Employment and Educational Settings: A Meta-Analysis*, 54 Personnel Psychol. 297 (2001) 24

Chase Sargent, *From Buddy to Boss: Effective Fire Service Leadership* (2006).....11, 23

Society for Industrial and Organizational Psychology, *Principles for the Validation and Use of Personnel Selection Procedures* (4th ed. 2003), *at* http://www.siop.org/_Principles/principles.pdf.....1, 2, 6, 7, 8, 9, 10, 13, 14, 15, 18, 27

Task Force on Assessment Center Guidelines, *Guidelines and Ethical Considerations for Assessment Center Operations*, 18 Pub. Personnel Mgmt. 457 (1989).....29, 30

Ian Taylor, *A Practical Guide to Assessment Centres and Selection Methods* (2007) 33

Michael A. Terpak, *Assessment Center: Strategy and Tactics* (2008) 15

George C. Thornton & Deborah E. Rupp, *Assessment Centers in Human Resource Management* (2006) 32

Carl F. Weaver, *Can Assessment Centers Eliminate Challenges to the Promotional Process?* (July 2000), *at* http://www.usfa.dhs.gov/pdf/efop/efo24862.pdf 30

Samuel J. Yeager, *Use of Assessment Centers by Metropolitan Fire Departments in North America*, 15 Pub. Personnel Mgmt. 51 (1986) 30

INTEREST OF *AMICI CURIAE*[1]

Amici are experts in the field of industrial-organizational psychology and are elected fellows of the Society for Industrial and Organizational Psychology ("SIOP"), the division of the American Psychological Association that is responsible for the establishment of scientific findings and generally accepted professional practices in the field of personnel selection. *Amici* also have extensive experience in the design and validation of promotional tests for emergency services departments, including fire and police departments across the country. *Amici* have an interest in ensuring the scientifically appropriate choice, development, evaluation, and use of personnel selection procedures.

Professor Herman Aguinis is the Mehalchin Term Professor of Management at the University of Colorado Denver Business School. In addition to his extensive scholarship and research on personnel selection, Professor Aguinis served on the Advisory Panel on the most recent revision of the *Principles for the Validation and Use of Personnel Selection Procedures* (4th ed. 2003) ("*Principles*"), *at* http://www.siop.org/_Principles/principles.pdf.

Professor Wayne Cascio holds the Robert H. Reynolds Chair in Global Leadership at the University of Colorado Denver Business School. He has published

---

[1] Pursuant to Supreme Court Rule 37.6, counsel for *amici* represents that it authored this brief in its entirety and that none of the parties or their counsel, nor any other person or entity other than *amici* or their counsel, made a monetary contribution intended to fund the preparation or submission of this brief. Counsel for *amici* also represents that all parties have consented to the filing of this brief in the form of blanket consent letters filed with the Clerk.

and testified extensively on issues relating to firefighter promotion. He served as President of SIOP from 1992 to 1993.

Professor Irwin Goldstein is currently Senior Vice Chancellor for Academic Affairs in the University System of Maryland. From 1991 to 2004, he served as Professor and Dean of the College of Behavioral and Social Sciences at the University of Maryland College Park. He has published extensively on validation, job selection, and training. He also served as President of SIOP from 1985 to 1986.

Dr. James Outtz, Ph.D., has more than 30 years' experience in the design and validation of personnel selection procedures. In 2002-2003, Dr. Outtz served on the Ad Hoc Committee that oversaw the 2003 revision of the *Principles*, the official policy statement of SIOP. Dr. Outtz also has served as a consultant for the City of Bridgeport, Connecticut, in designing the city's promotional examinations for fire lieutenant and captain positions.

Professor Sheldon Zedeck is Professor of Psychology at the University of California at Berkeley and Vice Provost for Academic Affairs and Faculty Welfare. Professor Zedeck's research and writing focuses on employment selection and validation models. Professor Zedeck served on the Ad Hoc Committee on the 2003 revision of the *Principles* and as President of SIOP from 1986 to 1987.

SUMMARY OF ARGUMENT

The City of New Haven ("City"), acting through its Civil Service Board ("Board"), reasonably declined to certify the results of the 2003 New Haven Fire Department ("NHFD") promotional examinations for captain and lieutenant because the validity of the tests could not have been substantiated under accepted scientific principles in the field of industrial-organizational ("I/O") psychology and applicable legal standards. Based on their expertise in the field of I/O psychology and their experience in employment test design, *amici* have identified at least four serious flaws in the tests that undermined their validity: (1) their admitted failure to measure critical qualifications for the job of a fire company officer; (2) the arbitrary, scientifically unsubstantiated weighting of the multiple-choice and oral components of the test battery; (3) the lack of input from local subject-matter experts regarding whether the tests matched the content of the jobs; and (4) use of strict rank-ordering without sufficient justification. Members of the Board thus reasonably concluded that it was unlikely, if not impossible, that the tests could be demonstrated to be valid.

Petitioners' claim that the decision not to certify the NHFD test results constituted a deviation from merit-based selection is inaccurate because of these clear and serious flaws in the design of the tests and the proposed use of the test scores. To the contrary, due to those flaws, which are apparent from the record below, there is no basis to conclude that certification of the test results would have led to the promotion of the most qualified candidates.

Moreover, several of the tests' flaws – namely, the unsubstantiated weighting of the test components

and use of strict rank-ordering – contributed to their adverse impact on racial subgroups, specifically African-American and Hispanic candidates. Thus, the tests not only failed to support an inference of superior job qualification from higher scores, but also simultaneously introduced a likely source of bias against minority candidates. In a predominantly minority city such as New Haven, bias against minority promotion exacerbates the public-safety risks of flawed tests by undermining the perception of fairness and cohesiveness among firefighters and by impairing the overall public effectiveness of the department. Thus, although the City had already approved and administered the test, the City appropriately concluded that costs to the NHFD and the local community would outweigh any potential benefit gained from certifying the tests.

The City could (and properly should) have adopted an alternative method of promotional selection to reduce the tests' adverse impact. At the very least, the City should have used scientifically substantiated weighting for the test components, which would likely have led to a reduced emphasis on the written component. Also, it could have discarded rank-ordering in favor of a "banding" approach, which treats candidates as equally qualified if their scores lie within a certain range reflecting the test's error of measurement. Rather than focus on the tests themselves, banding focuses on how test scores are used to make hiring decisions. Banding has been demonstrated, in some circumstances, to produce modest reductions in adverse impact without compromising the validity of the testing procedures in question. Moreover, the City could have adopted other options such as an "assessment center" that included behav-

ioral simulations of critical job components as part of the exams. Over the last 30 years, I/O psychology research has robustly confirmed that a properly validated assessment center can substantially reduce the adverse impact against minority candidates in the context of jobs such as firefighting.

In sum, given the flaws in the NHFD exams, which exacerbated the adverse impact on minority candidates, and given the availability of proven alternative selection methods, the City had reasonable, race-neutral grounds for deciding against certifying the results of the flawed tests. Indeed, under Title VII of the Civil Rights Act of 1964, it had *no choice* but to decline to certify the results. Petitioners' attempt to turn a decision *compelled by* Title VII into a violation of Title VII on the basis of mere insinuations about the Board's supposed racial biases turns the statute on its head and should be rejected.

ARGUMENT

I. THE 2003 EXAMINATIONS CONTAINED FATAL FLAWS THAT UNDERMINED THEIR ABILITY TO SELECT THE MOST QUALIFIED CANDIDATES

A critical and oft-repeated premise of petitioners' brief is that the 2003 NHFD examinations were "composed and validated" based on the *Uniform Guidelines on Employee Selection Procedures* ("*Uniform Guidelines*"), *see* 29 C.F.R. pt. 1607 (EEOC), and thus (1) actually "served their purpose of screening out the unqualified and identifying the most qualified," and (2) would have withstood scrutiny in a disparate-impact lawsuit brought by minority officer candidates. Pet. Br. 7, 35; *see also id.* at i (assuming in the question presented that the exams were "content-valid"). However, petitioners' premise lacks scientific

foundation. Under applicable I/O psychology principles and legal standards, there was no reasonable likelihood that the City could have demonstrated that the NHFD promotional examinations were valid.

A. Proper Validation Of Employment Tests According To Established Standards Is Essential To Ensuring Fair, Merit-Based Selection

A central objective of the field of I/O psychology is to develop generally accepted professional standards for the design and use of personnel-selection procedures, including employment tests, based on scientific data and analysis. The federal government has issued the *Uniform Guidelines*, which establish a federal standard for employment testing, *see* 29 C.F.R. § 1607.1(A), and "are intended to be consistent with generally accepted professional standards for evaluating standardized tests and other selection procedures," including the American Psychological Association's *Standards for Educational and Psychological Tests* ("APA *Standards*"), *id.* § 1607.5(C); *see Griggs v. Duke Power Co.*, 401 U.S. 424, 433-34 (1971) (holding that the *Uniform Guidelines* are "entitled to great deference"). The *Principles* are designed to be consistent with the APA *Standards* and represent the official policy statement of the Society for Industrial and Organizational Psychology ("SIOP"), a division of the APA, regarding current I/O psychology standards for personnel selection. *See Principles* at ii, 1.

"Validity is the most important consideration in developing and evaluating selection procedures." *Id.* at 4. "Validation" is the process of confirming that an employment test is "predictive of or significantly correlated with important elements of job perform-

ance." 29 C.F.R. § 1607.5(B).[2] In this case, for the reasons set forth below, at least four aspects of the NHFD promotional tests were flawed or arbitrary, and thus made it all but impossible for the City to show that the tests were valid.

The lack of evidence supporting the validity of the NHFD tests undermines their value as a selection tool. Proper validation of an employment test is critical to merit-based personnel selection because it ensures that there is a scientific basis for inferring that a higher test score corresponds to superior job skills or performance. *See Principles* at 4 (defining validity as "the degree to which accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of a test"); 29 C.F.R. § 1607.1(B). Moreover, proper validation promotes fairness and equal opportunity by ensuring that any disparate impact on subgroups is traceable to job requirements rather than contamination or bias in testing methodology. *See* 29 C.F.R. § 1607.3(A)

---

[2] The *Principles* and the *Uniform Guidelines* identify three types of evidence that support an inference of validity: "content," "criterion," and "construct." *See Principles* at 4-5; 29 C.F.R. § 1607.14(B)-(D). The three are not mutually exclusive categories, but rather refer to evidence supporting the inference that a test identifies those who are qualified to do the job. *See Principles* at 4. Content validity supports such an inference by showing that the test's content matches the essential content of the job, while criterion validity supports the inference by showing that test results successfully predict job performance. Construct validity is more abstract and is shown by evidence that the test measures the degree to which candidates have characteristics, or traits, that have been determined to lead to successful job performance. The flaws described in this brief undermine the content validity of the NHFD exams, which is the only evidence of validity asserted by petitioners and the only feasible type of validation in these circumstances.

(providing that a procedure that has an adverse impact "will be considered to be discriminatory . . . unless the procedure has been validated in accordance with these guidelines"); *Principles* at 7.

Validation is especially critical in the context of promotional exams for important public-safety leadership positions, such as fire company officers. Ensuring the selection of the most qualified fire officers saves lives. Accordingly, all state and local governments have a strong, race-neutral interest in declining to use promotional tests for fire officers that are shown to lack validity. Indeed, a legal regime in which state and local governments are hamstrung into implementing the results of such tests threatens the lives of the citizens they are committed to protect.[3] In this case, the City acted reasonably by declining to certify the NHFD promotional test results because the tests were fatally flawed.

B. Contrary To Petitioners' Premise, The 2003 NHFD Examinations Could Not Have Been Validated Under Established Principles

Petitioners' premise that the NHFD tests were properly validated rests on insufficient factual evidence, consisting merely of the fact that the test

---

[3] While critical, the validity of a test is only one of several issues that may legitimately be taken into account in making a decision on whether and how to use a test. The decision to use a particular test, or to use a test in a particular way, is made within a broader social and organizational context and appropriately takes into account possible "side effects" that could be costly or detrimental for the organization and the people served by the organization. *See* Herman Aguinis & Erika Harden, *Will Banding Benefit My Organization? An Application of Multi-Attribute Utility Analysis*, *in Test-Score Banding in Human Resource Selection* 193, 196-211 (Herman Aguinis ed., 2004).

designer, I/O Solutions ("IOS"), conducted what has been described as a job analysis, using "questionnaires, interviews, and ride-along exercises with incumbents to identify the importance and frequency of essential job tasks," and then had only one individual, a fire battalion chief in Georgia, review the tests. Pet. Br. 7; *see also id.* at 52. Petitioners also claim that the test designer provided "oral assurance of validity" and that the NHFD's Chief and Assistant Chief "thought the exams were fair and valid." *Id.* at 35.

Contrary to petitioners' assertions, conducting a job analysis – while in most cases necessary to a test's validity – is not alone sufficient to demonstrate validity. *See*, *e.g.*, 29 C.F.R. § 1607.14(B). Moreover, the *Uniform Guidelines* specifically reject the use of "casual reports of [a test's] validity," such as "testimonial statements and credentials of" IOS, and "non-empirical or anecdotal accounts" such as the comments of the NHFD's Chief and Assistant Chief. *Id.* § 1607.9. What the *Uniform Guidelines* and the *Principles* require is a rigorous analysis of the design and proposed use of the exam according to accepted principles of I/O psychology.[4]

Judged against the proper standards, there was no reasonable likelihood that the examinations administered by the NHFD could have been demonstrated to be valid, after the fact, according to generally accepted strategies for validation.[5]

---

[4] The fact that IOS was purportedly prepared to issue a validation study does not prove the tests were valid because the test-design process that the study would have described was fatally flawed.

[5] Petitioners' contention that the City avoided a full-fledged analysis of the validity of the examinations because it would

1. The Test Designer Conceded That the Exams Did Not Attempt To Measure Command Presence, a Critical Job Attribute

It is a fundamental precept of personnel selection that an employment test should be constructed to measure important knowledge, skills, abilities, and other personal characteristics ("KSAOs") needed for the job. *See* 29 C.F.R. §§ 1607.14(B)(3), 1607.14(C)(4). The omission from the testing domain of a KSAO that is an important job prerequisite – known in I/O psychology as "criterion deficiency" – vitiates the entire justification for the employment test, which is to select individuals accurately based on their capacity to perform the job in question. A test that makes no attempt to measure one or more critical KSAOs cannot be validated under established standards. *See*, *e.g.*, *Principles* at 23; *see also Firefighters Inst. for Racial Equality v. City of St. Louis*, 549 F.2d 506, 512 (8th Cir. 1977) ("*FIRE I*") (validity "requires that an important and distinguishing attribute be tested in some manner to find the best qualified applicants").

As the City's then-corporate counsel, Thomas Ude, recognized, the distinguishing feature of the job of a fire officer, as opposed to an entry-level firefighter, is responsibility for supervising and leading other firefighters in the line of duty. *See* JA138-39; *see also* Matthew Murtagh, *Fire Department Promotional Tests* 152 (1993) ("[C]ompany officers, lieutenants and

---

have been required to certify the test results is thus unsupportable. *See* U.S. Br. 19 ("An employer acts reasonably in not incurring [the burdens of a validity study] when it has significant questions concerning a test's job-relatedness or reasonably believes that better alternatives to the test may exist.").

captains, are the primary supervisors."); Anthony Kastros, *Mastering the Fire Service Assessment Center* 45 (2006). Leadership in emergency-response crises requires expertise in fire-management techniques and sound judgment about life-and-death decisions. Moreover, critically, it also requires a steady "presence of command" so that the unit will follow orders and respond correctly to fire conditions. *See*, *e.g.*, Richard Kolomay & Robert Hoff, *Firefighter Rescue & Survival* 5 (2003). Command presence requires an officer on the scene of a fire to act decisively, to communicate orders clearly and thoroughly to personnel on the scene, and to maintain a sense of confidence and calm even in the midst of intense anxiety, confusion, and panic. *See id.* at 5-13. Command presence generates respect for the officer among subordinates and is thus essential to order and discipline within the unit. *See* Murtagh at 152.

Simply put, command presence is a hallmark of a successful fire officer. *See*, *e.g.*, Chase Sargent, *From Buddy to Boss: Effective Fire Service Leadership* 21 (2006) ("No individual leader ever forgets the first time that their command presence was put to the test."). Virtually all studies of fire management emphasize that command presence is vital to the safety of firefighters at the scene and to the successful accomplishment of the firefighting mission and the safety of the public. *See*, *e.g.*, John F. Coleman, *Incident Management for the Street-Smart Fire Officer* 21-26 (2d ed. 2008); Kastros at 45; Vincent Dunn, *Command and Control of Fires and Emergencies* 1-6 (1999).

Here, the developer of the NHFD promotional exams, IOS, admitted that those exams were not designed to measure "command presence." Pet. App.

738a (Legel Dep.) ("Command presence usually doesn't come up as one of the skills and abilities we even try to assess.").[6] A high test score thus could not support an inference that the candidate would be a good commander in the line of duty; conversely, those candidates with strong command attributes were never given an opportunity to demonstrate them. Given the importance of command presence to the job of a fire officer, as the City recognized, this failure alone rendered the tests deficient. *See* JA139 (testimony of Mr. Ude that the "goal of the test is to decide who is going to be a good supervisor ultimately, not who is going to be a good test-taker").

In *FIRE I*, the Eighth Circuit recognized, in a similar situation, that the St. Louis fire captain's exam contained the "fatal flaw" of failing to test for "supervisory ability." 549 F.2d at 511. Because "supervisory ability" was a central requirement to the job of a fire captain, that failure precluded validation of the tests under the *Uniform Guidelines*. *Id.* The admitted failure of IOS to test command presence, a key attribute for any supervisory fire officer, would have led to the same result in this case.

2. The Weighting of the Multiple-Choice and Oral Interview Portions of the Exams Was Arbitrary and Could Not Be Validated

Even putting aside the failure of the NHFD exams to measure a critical aspect of the fire officer's responsibilities, the tests were seriously flawed for a second reason, stemming from the imposition of a

---

[6] However, IOS's representative, Chad Legel, conceded that "command presence" could be measured through the use of an assessment center. *See infra* p. 31.

predetermined 60/40 weighting for the written and oral interview components of the tests with no evidence that those weights matched the content of the jobs. Scientific principles of employment-test validation require not only that tests reflect the important KSAOs of the job, but also that the results from those tests be weighted in proportion to their job importance. Indeed, even assuming that a test measured the full range of important KSAOs, which the NHFD tests did not, *see supra* Part I.B.1, a test that gives inappropriate weight to some KSAOs over others could not have been shown accurately to select the candidates who are the most qualified for the job.[7] Numerous federal courts have likewise held that tests that measure relevant job skills without appropriate consideration for, and weighting of, their relative importance cannot properly be validated.[8]

---

[7] *See*, *e.g.*, *Principles* at 23-25 (explaining that selection procedures should adequately cover the requisite KSAOs and that the sample of KSAOs tested should be rationally related to the KSAOs needed to perform the work); Lynn Lyons Morris et al., *How to Measure Performance and Use Tests* 99 (1987) (explaining that a content-valid test should include a representative sample of categories of KSAOs and "give *emphasis* to each category according to its importance"); Robert D. Gatewood & Hubert S. Feild, *Human Resource Selection* 178-80 (5th ed. 2001) ("For our measures to have content validity, . . . their content [must] representatively sample[] the content of [the relevant performance] domains.").

[8] *See*, *e.g.*, *Isabel v. City of Memphis*, No. 01-2533 ML/BRE, 2003 WL 23849732, at *5 (W.D. Tenn. Feb. 21, 2003) ("[T]he test developer must demonstrate that those tests utilized in the selection system appropriately weigh the [KSAOs] to the same extent they are required on the job."), *aff'd*, 404 F.3d 404 (6th Cir. 2005); *Pina v. City of East Providence*, 492 F. Supp. 1240, 1246 (D.R.I. 1980) (invalidating a ranking system for firefighters that gave equal weight to written and physical components of the exam, on the ground that the physical component

The NHFD exams in this case failed that basic principle because the predetermined, arbitrary 60/40 weighting used to calculate the candidates' combined scores was in no way linked to the relative importance of "work behaviors, activities, and/or worker KSAOs," as required for validation. *Principles* at 25. The 60/40 weighting was determined in advance by the City's collective bargaining agreement with the local firefighters' union. *See* Pet. App. 606a (Legel Dep.). While it is not uncommon for municipal governments such as the City to enter into labor or other agreements that provide for a specified weighting of test components, such provisions undermine the validity of the resulting tests unless measures are taken by the test designer to account for the preset weighting.

Here, IOS concededly made no effort to establish that the 60/40 weighting was appropriate for the tests it designed. IOS should have used established methods to calculate whether, in light of the mandatory 60/40 weighting, the test components measured the job-relevant KSAOs in proportion to their relative importance. *See*, *e.g.*, Murtagh at 161; Gatewood & Feild at 178. IOS apparently did not do so; instead, it merely assessed whether the test questions were related to relevant aspects of the job, with no regard to whether the items included on the test proportionally measured the critical aspects of the overall job.

"appear[ed] to have a greater relationship to job performance as a firefighter"); *Jones v. New York City Human Res. Admin.*, 391 F. Supp. 1064, 1079 n.15, 1081 (S.D.N.Y. 1975) (calling it a "serious defect" not to examine "how or why the skills in the test plan were weighted as they were," resulting in a test that could not be validated; for example, numerous questions related to a skill, supervision, that only 60%-65% of employees in the position actually performed), *aff'd*, 528 F.2d 696 (2d Cir. 1976).

See Pet. App. 634a (Legel Dep.). IOS's failure to take that step resulted in tests that, absent sheer luck, could not have resulted in adequate validity evidence under the *Principles* or the *Uniform Guidelines*.

Moreover, there is no indication that the 60/40 weighting at issue in this case, which gave predominance to the multiple-choice component of the exams, was appropriate for the relevant job. It is well-recognized by I/O psychologists and firefighters alike that written, pencil-and-paper tests, while able to measure certain cognitive abilities (*e.g.*, reading and memorization) and factual knowledge, do not measure other skills and abilities critical to being an effective fire officer as well as alternative methods of testing do. *See*, *e.g.*, Michael A. Terpak, *Assessment Center: Strategy and Tactics* 1 (2008) (multiple-choice exams are "known to be poor at measuring the knowledge and abilities of the candidate, most notably that of a fire officer"); Int'l Ass'n of Fire Chiefs et al., *Fire Officer: Principles and Practice* 28 (2006) (describing the criticism of written tests as producing firefighters who are "[b]ook smart, street dumb"); *see also* David L. Bullins, *Leading in the Gray Area*, Fire Chief (Aug. 10, 2006) ("Good leadership is not a matter of decisions made in black and white; it is a matter of the decisions that must be made in shades of gray."), *at* http://firechief.com/management/bullins_gray08102006/index.html. Although a written component often properly comprises part of the overall assessment procedure for fire officers, a weighting of 60% is significantly above what would be expected given the requirements of the positions. *See* Phillip E. Lowry, *A Survey of the Assessment Center Process in the Public Sector*, 25 Pub. Personnel Mgmt. 307, 309 (1996) (survey finding that the median weight given

to written portion of test for fire and police departments was 30%); *infra* p. 26 (describing weights used by neighboring Bridgeport).

The *Uniform Guidelines* and the federal courts have similarly recognized that written tests do not correspond well to the skills and abilities actually required for the job of a fire officer and are thus poor predictors of which candidates will make successful fire lieutenants and captains. The EEOC's interpretive guidance on the *Uniform Guidelines*[9] states that "[p]aper-and-pencil tests of . . . ability to function properly under danger (*e.g.*, firefighters) generally are not close enough approximations of work behaviors to show content validity." *Questions and Answers* No. 78, 44 Fed. Reg. at 12,007.

The Eighth and Eleventh Circuits have reached the same conclusion. In *FIRE II*, the Eighth Circuit rejected the validity of a multiple-choice test for promotion to fire captain, on the ground that "[t]he captain's job does not depend on the efficient exercise of extensive reading or writing skills, the comprehension of the peculiar logic of multiple choice questions, or excellence in any of the other skills associated with outstanding performance on a written multiple choice test." 616 F.2d at 357. "'Where the content

---

[9] The EEOC's interpretive "questions and answers" were adopted by the four agencies that promulgated the *Uniform Guidelines* in order to "interpret and clarify, but not to modify," those *Guidelines*. *Adoption of Questions and Answers To Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures*, 44 Fed. Reg. 11,996, 11,996 (1979) ("*Questions and Answers*"). Like the *Uniform Guidelines* themselves, the agency interpretations have been given great deference by the courts. *See*, *e.g.*, *Firefighters Inst. for Racial Equality v. City of St. Louis*, 616 F.2d 350, 358 n.15 (8th Cir. 1980) ("*FIRE II*").

and context of the selection procedures are unlike those of the job, as, for example, in many paper-and-pencil job knowledge tests, it is difficult to infer an association between levels of performance on the procedure and on the job.'" *Id.* at 358 (quoting *Questions and Answers* No. 62, 44 Fed. Reg. at 12,005). Accordingly, "[b]ecause of the dissimilarity between the work situation and the multiple choice procedure," the court found that "greater evidence of validity [wa]s required." *Id.* at 357.

In *Nash v. Consolidated City of Jacksonville*, 837 F.2d 1534 (11th Cir. 1988), *vacated and remanded*, 490 U.S. 1103 (1989), *opinion reinstated on remand*, 905 F.2d 355 (11th Cir. 1990), the Eleventh Circuit likewise rejected the use of a written test to determine eligibility for promotion to the position of fire lieutenant. The court rejected the use of the test even though the test questions "never made their way into evidence" and even though the expert who was challenging the use of the test on behalf of the firefighter had never seen the questions. *Id.* at 1536. As the court explained, "[a]n officer's job in a fire department involves 'complex behaviors, good interpersonal skills, the ability to make decisions under tremendous pressure, and a host of other abilities – none of which is easily measured by a written, multiple choice test.'" *Id.* at 1538 (quoting *FIRE II*, 616 F.2d at 359).

IOS exacerbated the problem of imbalance in its response to another predetermined feature of the NHFD exams – the 70% cutoff score mandated by the City's civil service rules. Like the 60/40 weighting, the 70% cutoff score was arbitrary and not scientifically validated. *See* Pet. App. 697a-698a (concession by Mr. Legel that IOS was unable to validate the

70% cutoff score).[10] IOS not only "went ahead and used [the] seventy percent," but also decided to make the written component of the test "more difficult" in an effort to screen out "a fair amount more number [sic] of people . . . than what other tests have done in the past." *Id.* at 698a-699a (Legel Dep.). Not only did this admittedly worsen the adverse impact of the tests on minority candidates, *see infra* Part II.A, but it also skewed the focus of the test even more heavily in the direction of the limited and more attenuated set of knowledge and abilities that are measured by a multiple-choice test, by giving that component unjustifiably greater weight in the composite scores. That, in turn, further reduced the likelihood that the exams could have been shown to be valid.[11]

---

[10] Arbitrary cutoff scores alone can undermine a test's validity. *See*, *e.g.*, *Isabel v. City of Memphis*, 404 F.3d 404, 413 (6th Cir. 2005) (stating that, "[t]o validate a cutoff score, the inference must be drawn that the cutoff score measures minimal qualifications"); *accord Chisholm v. United States Postal Serv.*, 516 F. Supp. 810, 832, 838 (W.D.N.C. 1980), *aff'd in relevant part*, 665 F.2d 482 (4th Cir. 1981). The *Uniform Guidelines* and the *Principles* clearly require cutoff scores, if they are used, to be based on scientifically accepted principles. *See* 29 C.F.R. § 1607.5(H) ("Where cutoff scores are used, they should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force."); *Principles* at 47 (explaining that "[p]rofessional judgment is necessary in setting any cutoff score" in light of factors including "the [KSAOs] required by the work"); *see also* Wayne F. Cascio & Herman Aguinis, *Test Development and Use: New Twists on Old Questions*, 44 Human Res. Mgmt. 219, 227 (2005) (discussing appropriate process for calibrating cutoff score to minimum proficiency for the job).

[11] In fact, IOS acknowledged that even the oral portion of the test was designed, at least "to a small degree," to test factual knowledge, thus further skewing the balance of the test. Pet. App. 709a (Legel Dep.).

Under established principles in the field of I/O psychology and longstanding legal authorities, the NHFD exams were deficient because of IOS's failure to substantiate the predetermined 60/40 weighting before administering the test and because of the resulting overemphasis given to the written, multiple-choice component of the exams, which has been demonstrated to be a relatively poor method for measuring whether a candidate has the KSAOs needed to be a fire officer.

3. Flaws in the Exam-Development Process Contributed to the Lack of Validity Evidence of the NHFD Tests

The process used to develop and finalize the tests further undermined the tests' validity as a method for identifying the individuals best suited for promotion. IOS personnel wrote the test questions based on the information developed from job analysis questionnaires given to incumbent New Haven fire officers and "national texts" on firefighting. C.A. App. 478 (Legel Dep.). However, IOS personnel were not themselves subject-matter experts on the job of a fire company officer, nor were the "national texts" they used tailored to the NHFD's specific practices or local conditions in New Haven. *See id.* ("So depending on the way that those [New Haven] City employees are trained to do their specific job, it may not always jibe with the way the textbook says to do it."); *see also* Pet. App. 520a-521a ("Fire fighting is different on the East Coast than it is on [sic] West Coast or in the Midwest.").

Accordingly, as IOS acknowledged, "[s]tandard practice" in the field required that the tests be reviewed by "a *panel* of subject matter experts internal to New Haven, for instance, incumbent lieutenants,

captains, battalion chiefs, [assistant] chiefs, and the like to actually gain [sic] their opinion about how relevant the items were and whether or not they were consistent with best practice in New Haven." *Id.* at 635a (Legel Dep.) (emphasis added). Review by multiple persons with specific expertise about the NHFD was, as IOS recognized, important to verify that the questions accurately reflected important KSAOs of the job and, especially, local differences between NHFD's practices and procedures and national firefighting standards. *See* Wayne F. Cascio & Herman Aguinis, *Applied Psychology in Human Resource Management* 158-59 (6th ed. 2005) (documenting the need for subject-matter experts to "confirm[] the fairness of sampling and scoring procedures" and to evaluate "overlap between the test and the job performance domain"); Irwin L. Goldstein et al., *An Exploration of the Job Analysis–Content Validity Process*, *in Personnel Selection in Organizations* 3, 20-21 (Neil Schmitt & Walter C. Borman eds., 1993); Int'l Ass'n of Fire Chiefs et al., *Fundamentals of Fire Fighter Skills* 103, 431, 663 (2004) (emphasizing that firefighters need to become "intimately familiar" with local procedures and local differences affecting firefighting such as architectural styles).

Rather than follow this admittedly standard procedure, IOS hired a single individual, a battalion chief in a fire department in Georgia, to review the tests for the job-relatedness of their content. *See* Pet. App. 635a-636a (Legel Dep.). Unsurprisingly, due to the failure to conduct a proper review by multiple subject-matter experts on local practice, IOS admitted that some of the items on the tests were "irrelevant for the City because you're testing them on a knowledge base that while supported by a national textbook,

wouldn't be supported by their own standard operating procedures." C.A. App. 482 (Legel Dep.). For example, the lieutenants' test included a question from a New York-based textbook about whether fire equipment should be parked uptown, downtown, or underground when arriving at a fire. JA48. The question was meaningless because New Haven has no "uptown" or "downtown."

By IOS's admission, and under applicable I/O psychology standards, review of the test items by local subject-matter experts was critical to ensuring that the test components corresponded to the important job KSAOs. The failure to do so further undermined the validity of the NHFD exams as indicators of which candidates would have made successful NHFD fire lieutenants or captains.

4. The NHFD Tests Could Not Have Been Validated for Strict Rank-Ordering of Candidates

Under accepted standards, not only must an exam's content be properly validated, but the use of the scores also must be scientifically justified. As the *Uniform Guidelines* state, "the use of a selection procedure on a pass/fail (screening) basis may be insufficient to support the use of the same procedure on a ranking basis under these guidelines." 29 C.F.R. § 1607.5(G). Under the *Uniform Guidelines*, a strict rank-ordering system such as the one imposed by the City – *i.e.*, treating a candidate as "better qualified" based on even a slight incremental difference in score – is only appropriate upon a scientific showing "that a higher score on a content valid selection procedure is likely to result in better job performance." *Id.* § 1607.14(C)(9). As the Second Circuit held in *Guardians Association of New York City Police Department*

v. Civil Service Commission, 630 F.2d 79 (2d Cir. 1980), "[p]ermissible use of rank-ordering requires a demonstration of such substantial test validity that it is reasonable to expect one- or two-point differences in scores to reflect differences in job performance." *Id.* at 100-01 (rejecting the validity of rank-ordering); *see also FIRE II*, 616 F.2d at 358.

In this case, the NHFD tests could not have supported the use of a strict rank-ordering procedure for promotional selection. Indeed, the tests were designed and administered at a time when New Haven's "Rule of Three" had been interpreted to permit rounding of scores to the nearest integer, rather than strict rank-ordering based on differences of fractions of a point. *See* C.A. App. 1701; *Kelly v. City of New Haven*, 881 A.2d 978, 993-94 (Conn. 2005). Use of strict rank-ordering for a test absent evidence demonstrating that it was valid for that purpose cannot be justified. *See Pina*, 492 F. Supp. at 1246 (invalidating test where "[t]here [wa]s no evidence which even remotely suggest[ed] that the order of ranking establishe[d] that any applicant [wa]s better qualified than any other").

Moreover, as explained above, the serious flaws in the NHFD tests severely undermined the overall validity of the exams and certainly foreclosed any conclusion that the exams were of such "substantial . . . validity" as to justify the additional step of making promotional decisions strictly based on small score differences. *Guardians Ass'n*, 630 F.2d at 100-01. Making fine judgments based on small differences on fundamentally flawed tests is scientifically unsupportable. *See*, *e.g.*, Aguinis & Harden at 193. "[U]se of an exam to rank applicants, when the exam cannot predict applicants' relative merits, offers

nothing but a false sense of assurance based on a misplaced belief that *some* criterion – no matter how arbitrary – is better than none." *Ensley Branch, NAACP v. Seibels*, 31 F.3d 1548, 1574 (11th Cir. 1994).

Tests that transform differences that are as likely to be a product of measurement error or flawed test design as they are a reflection of superior qualifications create nothing but the illusion of meritocracy. That illusion creates not only a false sense of individual entitlement to jobs and promotions, but also a real public danger in the context of positions such as fire and police officers. When the safety and lives of citizens are at stake, it is particularly critical for public employers to have the leeway to ensure that the tests they deploy accurately identify those candidates who are most qualified for these important jobs.

II. THE FLAWS IN THE NHFD PROMOTIONAL EXAMS EXACERBATED THEIR ADVERSE IMPACT ON MINORITY CANDIDATES

Unjustified exclusion of minority candidates through scientifically flawed testing procedures has significant social costs. Especially in a city like New Haven, racial diversity has significant benefits to the ability of the public sector to provide needed services to the community and to protect the public safety. *See*, *e.g.*, Wayne F. Cascio et al., *Social and Technical Issues in Staffing Decisions, in Test Score Banding, in Human Resource Selection* 7, 9 (Herman Aguinis ed., 2004). An all-white officer corps in the NHFD will be less effective than one that is more racially diverse. *See id.*; *see also* Sargent at 188 (noting that having a Hispanic firefighter fluent in Spanish "can be a life saver").

In this case, the flaws in the NHFD promotional exams not only undermined their validity, but also unjustifiably increased their adverse impact on minority candidates. In particular, two features of the tests contributed to the conceded adverse impact on African-American and Hispanic examinees. Tests that eliminated these features were available to the City as "less discriminatory alternatives" under Title VII.

A. Overweighting Of The Written, Multiple-Choice Portion Of The Exams Increased The Adverse Impact On Minority Candidates

It is well-established that minority candidates fare less well than their Caucasian counterparts on standardized written examinations, and especially multiple-choice (as opposed to "write-in") tests. *See*, *e.g.*, Winfred Arthur Jr. et al., *Multiple-Choice and Constructed Response Tests of Ability*, 55 Personnel Psychol. 985, 986 (2002); Philip L. Roth et al., *Ethnic Group Differences in Cognitive Ability in Employment and Educational Settings: A Meta-Analysis*, 54 Personnel Psychol. 297 (2001). Although the causes for that widely recognized discrepancy are not fully understood, certain features of the multiple-choice format have been recognized to contribute to adverse impact.

First, "[t]o the extent that [the exam's] reading demands are not concomitant with job demands and/or performance, then any variance associated with reading demands and comprehension is considered to be error variance." Arthur et al., 55 Personnel Psychol. at 991. Some studies suggest disparities among racial subgroups in reading comprehension, such that using written questions and answers as

the sole or predominant medium for testing increases adverse impact. *See id.*; James L. Outtz & Daniel A. Newman, *A Theory of Adverse Impact* 12-13, 68 (manuscript on file with author) (forthcoming in *Adverse Impact: Implications for Organizational Staffing and High Stakes Selection*, 2009). Moreover, studies suggest that racial minorities are less "test wise" than white test-takers, and it is "widely recognized that performance on multiple-choice tests is susceptible to specific test-taking strategies or test-wiseness." Arthur et al., 55 Personnel Psychol. at 991-92. Finally, studies have found that a test-taker's unfavorable view of a test's validity negatively influences performance, and some evidence indicates that minority test-takers generally have a less favorable view of traditional written tests. *See id.* at 992.

Regardless of the exact cause of the disparity, it is clear that the use of written, multiple-choice tests beyond what is justified by the demands of a particular job has the effect of disproportionately excluding minority candidates without any corresponding increase in job performance. *See*, *e.g.*, Outtz & Newman at 33. As set forth above, the NHFD's 60/40 weighting was arbitrary and put more emphasis on the written, multiple-choice examination than science and experience have shown to be warranted for the job of a fire officer. Likewise, the response of IOS to the 70% cutoff score contributed to the adverse impact of the exams. By IOS's own admission, arbitrarily making the written portion of the tests "more difficult" further exaggerated the importance of the written component and thereby contributed to the exclusion of African-American and Hispanic candi-

dates from the promotional ranks. Pet. App. 698a-699a (Legel Dep.).

Changing the weighting of the exams to more accurately reflect the content of the job almost certainly would have reduced their adverse impact by reducing the weight of the written component, and thus constituted a "less discriminatory alternative" that the City would have been obligated to use under Title VII. Had the City given a 30% weighting to the written component of the examination, more in line with the nationwide norm, *see supra* pp. 15-16, the tests would have had a significantly lower adverse impact on minority candidates. *See* Resp. Br. 33 ("[I]f the tests were weighted 70%/30% oral/written, then two African-Americans would have been considered for lieutenant positions and one for a captain position."). Indeed, 20 miles down the coast from New Haven, the fire department in Bridgeport, Connecticut, has administered tests with less weight given to the written component (25% for lieutenants and 33% for captains) and achieved a significant reduction in adverse impact relative to the NHFD exam results. *See* JA64-66.[12]

B. Selection Of Candidates In Strict Rank Order Also Contributed To Adverse Impact

As discussed above, the NHFD tests were improperly weighted toward the written component, which tested certain KSAOs (*e.g.*, reading, memorization, and factual knowledge) in disproportion to their im-

---

[12] In 2005, the Bridgeport lieutenant's exam consisted of written and oral components, weighted 25% and 75%, respectively; the weights on the captain's exam were 33% for the written component, 62% for the oral component, and 5% for seniority.

portance relative to other important skills and abilities, including "command presence," which was not measured at all. Moreover, the tests unjustifiably employed a strict rank-ordering system that differentiated among candidates based on small score differences that had not been scientifically demonstrated to be meaningful. The combination of imbalanced weighting toward KSAOs that disproportionately disfavor minority candidates and the selection of candidates based strictly on rank order cemented the disproportionate rejection of minority candidates for promotion.

An alternative to strict rank-ordering would have been a "banding" scoring system. In brief, banding involves use of statistical analysis of the amount of error in the test scores to create "bands" of scores, the lowest of which is considered to be sufficiently similar to the highest to warrant equal consideration within that band. Cascio et al. at 10; *see also Principles* at 48 (bands "take into account the imprecision of selection procedure scores and their inferences"). After the width of the band is established, based on a statistical analysis of the reliability of measurement, the user can either establish "fixed" bands, in which the test user considers everyone within the top band before considering anyone from the next band, or "sliding" bands, which allows the band to "slide" down the list once higher scorers are either chosen or rejected. *See* Cascio et al. at 10-11.

The federal courts have recognized banding as "a universal and normally an unquestioned method of simplifying scoring by eliminating meaningless gradations" between candidates whose scores differ by less than the degree of measurement error. *Chicago Firefighters Local 2 v. City of Chicago*, 249 F.3d 649,

656 (7th Cir. 2001); *see also*, *e.g.*, *Biondo v. City of Chicago*, 382 F.3d 680, 684 (7th Cir. 2004) (banding "respect[s] the limits of [an] exam's accuracy"). In *amici*'s view, a banding approach would have been a viable method to reduce the adverse impact of the NHFD tests. However, given that the rankings themselves were a result of flawed tests, banding alone would not have been sufficient to achieve the objective of selecting the most qualified individuals for the job. *See Bridgeport Guardians, Inc. v. City of Bridgeport*, 933 F.2d 1140, 1147 (2d Cir. 1991) ("The ranking of the candidates was itself the result of the disparate impact of the examination.").

III. CURRENT I/O PSYCHOLOGY RESEARCH SUPPORTS THE USE OF PROMOTIONAL ASSESSMENT CENTERS AS A VALID AND LESS DISCRIMINATORY ALTERNATIVE TO TRADITIONAL TESTING METHODS

The evidence in the record clearly demonstrates that the NHFD exams suffered from fatal design defects that undermined their validity and unjustifiably excluded a disproportionate number of minority candidates. That alone left the City no choice but to decline to certify the exams. In addition, the City reasonably concluded that certification of the tests could not be justified given the existence of alternative methods of selection. One alternative before the City was the assessment center, which, if designed properly, would measure a broader range of KSAOs and also be less discriminatory. *See*, *e.g.*, JA96 (statement of Dr. Christopher Hornick to the Board that assessment centers are "much more valid in terms of identifying the best potential supervisors"); Pet. App. 739a (Legel Dep.).

A. History Of The Assessment Center Model

From the 1950s to the 1980s, multiple-choice tests were generally the only procedure used for promotional selection in U.S. fire departments. *See* Int'l Ass'n of Fire Chiefs et al., *Fire Officer: Principles and Practice* at 28. Such tests were prevalent because they were easy and inexpensive to administer, and seemingly "objective." However, for the reasons discussed above, such tests had the side-effect of excluding a disproportionate number of minority candidates from consideration. Beginning in the 1970s, spurred in part by the passage of Title VII (1964), the development of the *Uniform Guidelines* (1966), and this Court's decision in *Griggs* (1971), employers increasingly began using an alternative selection method known as the assessment center. *See* James R. Huck & Douglas W. Bray, *Management Assessment Center Evaluations and Subsequent Job Performance of White and Black Females*, 29 Personnel Psychol. 13, 13-14 (1976).

An assessment center is a form of standardized evaluation that seeks to test multiple dimensions of job qualification through observation of job-related exercises and other assessment techniques. *See generally* Task Force on Assessment Center Guidelines, *Guidelines and Ethical Considerations for Assessment Center Operations*, 18 Pub. Personnel Mgmt. 457, 460-64 (1989) (defining an assessment center). Unlike multiple-choice exams, which evaluate KSAOs through a single, written medium, assessment centers employ multiple methods, including, prominently, job simulations, all of which are designed to permit more direct assessment of ability to do the job. *See id.* at 461-62. Candidates' performance on the simulation exercises is rated by multiple subject-matter

experts. *See id.* at 462. By observing how a participant handles the problems and challenges of the target job (as simulated in the exercises), assessors develop a valid picture of how that person would perform in the target position. *See* Charles D. Hale, *The Assessment Center Handbook for Police and Fire Personnel* 16-52 (2d ed. 2004) (describing typical exercises).

B. Assessment Centers Have Demonstrated Validity In The Context Of Firefighter Promotion

Since the 1970s, the use of assessment centers for employee selection has increased rapidly, both in the United States and elsewhere, and in firefighter promotion in particular. By 1986, 44% of fire departments surveyed used assessment centers in making promotion decisions.[13] More recent surveys indicate a usage rate of between 60% and 70%.[14]

Like any testing method, an assessment center must be properly constructed so that, for example, it measures important KSAOs of the relevant job. After more than 30 years of use and research, however, substantial agreement exists among I/O psychologists that properly designed assessment centers are better predictors of job performance than other

---

[13] *See* Samuel J. Yeager, *Use of Assessment Centers by Metropolitan Fire Departments in North America*, 15 Pub. Personnel Mgmt. 51, 52-53 (1986); *accord* Miguel Roig & Maryellen Reardon, *A Performance Standard for Promotions*, 141 Fire Engineering 49, 49 (1988).

[14] *See* Lowry, 25 Pub. Personnel Mgmt. at 310; Carl F. Weaver, *Can Assessment Centers Eliminate Challenges to the Promotional Process?* 13 (July 2000) (unpublished monograph), *at* http://www.usfa.dhs.gov/pdf/efop/efo24862.pdf.

forms of promotional testing.[15] Today, because of numerous studies supporting the conclusion, "the predictive validity of [assessment centers] is now largely assumed." Walter C. Borman et al., *Personnel Selection*, 48 Ann. Rev. Psychol. 299, 313 (1997). Properly designed assessment centers have incremental predictive validity over cognitive tests "because occupational success is not only a function of a person's cognitive abilities, but also the *manifestation* of those abilities in concrete observable behavior." Diana E. Krause et al., *Incremental Validity of Assessment Center Ratings Over Cognitive Ability Tests: A Study at the Executive Management Level*, 14 Int'l J. Selection & Assessment 360, 362 (2006).

As reflected by their widespread usage by fire departments across the country, assessment centers are especially appropriate in the context of firefighter promotion. Because they use multiple methods of assessment, assessment centers are able to measure a wider range of skills, including critical skills such as leadership capacity, problem-solving, and "command presence." IOS's representative, Chad Legel, admitted that the NHFD tests failed to test for "command presence," and he further acknowledged that the City "would probably be better off with an assessment center if you cared to measure that." Pet. App. 738a (Legel Dep.); *see also* Krause et al., 14 Int'l J. Selection & Assessment at 362 (agreeing that leadership ability is likely better assessed through an

---

[15] *See, e.g.*, Chaitra M. Hardison & Paul R. Sackett, *Assessment Center Criterion Related Validity: A Meta-Analytic Update* 14-20 (2004) (unpublished manuscript); Winfred Arthur Jr. et al., *A Meta-Analysis of the Criterion-Related Validity of Assessment Center Dimensions*, 56 Personnel Psychol. 125, 145-46 (2003); Barbara B. Gaugler et al., *Meta-Analysis of Assessment Center Validity*, 72 J. Applied Psychol. 493, 503 (1987).

assessment center than an oral interview); Gaugler et al., 72 J. Applied Psychol. at 493 ("assessment centers are most frequently used for assessing managers").

In short, the "state of the art" in the field of promotional testing for firefighters and the "state of the science" in I/O psychology have evolved beyond the outdated methods of testing used by the NHFD. Instead, as the City was told by Dr. Hornick, *see* JA96, there is now substantial agreement that a professionally validated assessment center represents a more effective method of selecting the most qualified fire officers.

C. Assessment Centers Have Been Proven To Reduce Adverse Impact On Minorities

It is equally well-recognized in the research literature that assessment centers reduce adverse impact on racial minorities as compared to traditional standardized tests. *See*, *e.g.*, George C. Thornton & Deborah E. Rupp, *Assessment Centers in Human Resource Management* 231 (2006). "Additional research has demonstrated that adverse impact is less of a problem in an assessment center as compared to an aptitude test designed to assess cognitive abilities that are important for the successful performance of work behaviors in professional occupations." Cascio & Aguinis, *Applied Psychology in Human Resource Management* at 372-73.

Those scientific studies also have been borne out by experience. An analysis of fire-personnel selection in St. Louis in the 15 years after the *FIRE II* decision found that the institution of an assessment center selection method "achieved considerable success at minimizing adverse impact against black candidates." Gary M. Gebhart et al., *Fire Service Testing*

in a Litigious Environment: A Case History, 27 Pub. Personnel Mgmt. 447, 453 (1998).

In sum, assessment centers are now a prevalent feature of firefighter promotional tests across the nation. Numerous resources exist for employers wishing to incorporate assessment centers into their selection procedures in accordance with accepted scientific principles.[16] The availability of the assessment center as an equally valid, less discriminatory alternative provides yet another justification for the City's decision not to certify the results of the NHFD promotional exams. Indeed, under Title VII, it compelled that decision.

* * * * *

To place this case in overall perspective, petitioners' lawsuit seeks to compel the City to certify the results of tests that suffered from glaring flaws undermining their validity, had an admitted adverse impact on racial minorities, and could have been replaced by readily available, equally or more valid, and less discriminatory alternatives. From the standpoint of accepted I/O psychology principles, there is no justification for certifying the results of such tests because there is no evidence they selected the most qualified candidates, and they systematically excluded minority candidates. Under established legal principles, moreover, certification would have resulted in a violation of Title VII, and the City's decision was thus compelled by law. Petitioners' challenge to the City's decision must therefore fail.

---

[16] *See generally*, *e.g.*, Ian Taylor, *A Practical Guide to Assessment Centres and Selection Methods* (2007); Hale, *supra*.

CONCLUSION

The judgment of the court of appeals should be affirmed.

Respectfully submitted,

DAVID C. FREDERICK
Counsel of Record
DEREK T. HO
BARRETT C. HESTER
JENNIFER L. PERESIE
KELLOGG, HUBER, HANSEN,
TODD, EVANS & FIGEL, P.L.L.C.
1615 M Street, N.W., Suite 400
Washington, D.C. 20036
(202) 326-7900

Counsel for Industrial-Organizational Psychologists

March 25, 2009